Protein Structure Prediction and Potential Energy Landscape Analysis using Continuous Global Minimization

KEN A. DILL
Department of Pharmaceutical Chemistry, University of California at San Francisco, San Francisco, CA 94118 <dill@maxwell.ucsf.edu>

ANDREW T. PHILLIPS
Computer Science Department, United States Naval Academy, Annapolis, MD 21402 <phillips@haney.scs.usna.navy.mil>

J. BEN ROSEN
*Computer Science and Engineering Department, University of California at San Diego, San Diego, CA 92093 <jbrosen@cs.ucsd.edu>

INTRODUCTION

Proteins require specific three-dimensional conformations to function properly. These "native" conformations result primarily from intramolecular interactions between the atoms in the macromolecule, and also intermolecular interactions between the macromolecule and the surrounding solvent. Although the folding process can be quite complex, the instructions guiding this process are specified by the one-dimensional primary sequence of the protein or nucleic acid: external factors, such as helper (chaperone) proteins, present at the time of folding have no effect on the final state of the protein. Many denatured proteins spontaneously refold into functional conformations once denaturing conditions are removed. Indeed, the existence of a *unique* native conformation, in which residues distant in sequence but close in proximity exhibit a densely packed hydrophobic core, suggests that this three-dimensional structure is largely encoded within the sequential arrangement of these specific amino acids. In any case, the native structure is often the conformation at the global minimum energy (see [1]).

In addition to the unique native (minimum energy) structure, other less stable structures exist as well, each with a corresponding potential energy. These structures, in conjunction with the native structure, make up an energy landscape that can be used to characterize various properties of the protein.

Over 20 years of research into this "protein folding problem" has resulted in numerous important algorithms that aim to predict native three-dimensional protein structures (see [2], [3], [4], [8], [9], [10], [12], [14], [15], [18], [19], [20], [21], and [22]). Such methods assume that the native structure is a balance of various interactions. These methods invariably use some form of energy minimization technique, such as simulated annealing or genetic algorithms, rather than the computationally more efficient continuous minimization techniques. Each such method correctly predicts a few protein structures, but misses many others, and the process is both slow and limited to structures of relatively small size.

The purpose of this paper is to show how one can apply more efficient continuous minimization techniques to the energy minimization problem by using an accurate continuous approximation to the discrete information provided for known protein structures. In addition, we will show how the results of one particular computational method for protein structure prediction (the CGU algorithm), which is based on this continuous minimization technique, can be used both to accurately determine the global minimum of potential energy function and also to offer a quantitative analysis of *all* of the local (and global) minima on the energy landscape.

The CGU method has been extensively tested on a variety of computational platforms including the Intel Paragon, Cray T3D, an 8 workstation Dec Alpha cluster, and a heterogeneous network of 13 Sun SparcStations and 7 SGI Indys. Protein structures with as many as 46 residues have been computed in under 40 hours on the 20 workstation heterogeneous network.

THE POLYPEPTIDE MODEL AND POTENTIAL ENERGY FUNCTION

Since computational search methods are not yet fast enough to find global optima in real-space representations using accurate all-atom models and potential functions, a practical conformational search strategy requires both a simplified, yet sufficiently realistic, molecular model with an associated potential energy function representing the dominant forces involved in protein folding. We also need a global optimization method which takes full advantage of any special properties of this kind of energy function.

Each residue in the primary sequence of a protein is characterized by its backbone components NH-C_αH-C′O and one of 20 possible amino acid sidechains attached to the central C_α atom. The three-dimensional structure of macromolecules is determined by internal molecular coordinates consisting of bond lengths *l* (defined by every pair of consecutive backbone atoms), bond angles θ (defined by every three consecutive backbone atoms), and the backbone dihedral angles φ, ψ, and ω, where φ gives the position of C′ relative to the previous three consecutive backbone atoms C′-N-C_α, ψ gives the position of N relative to the previous three consecutive backbone atoms N-C_α-C′, and ω gives the position of C_α relative to the previous three consecutive backbone atoms C_α-C′-N. Figure 1 illustrates this model.

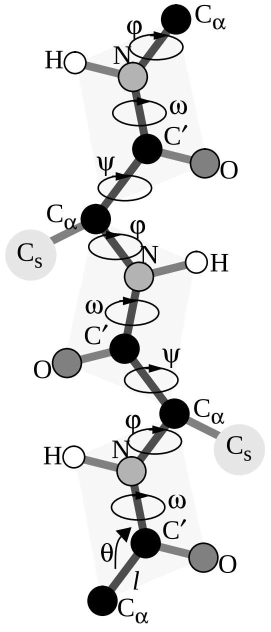

Figure 1 Simple Polypeptide Model

Fortunately, these 9*n*-6 parameters (for an *n*-residue structure) do not all vary independently. In fact, some of these (7*n*-4 of them) are regarded as fixed since they are found to vary within only a very small neighborhood of an experimentally determined value. Among these are the 3*n*-1 backbone bond lengths *l* between the pairs of consecutive atoms N-C′, C′-C_α, and C_α-N. Also, the 3*n*-2 backbone bond angles θ defined by N-C_α-C′, C_α-C′–N, and C′-N-C_α are also fixed at their ideal values. Finally, the *n*-1 peptide bond dihedral angles ω are fixed in the trans (180˚) conformation. This leaves only the *n*-1 backbone dihedral angle pairs (φ,ψ) in the reduced representation model. These also are not completely independent; they are severely constrained by known chemical data (the Ramachandran plot) for each of the 20 amino acid residues. Furthermore, since the atoms from one C_α to the next C_α along the backbone can be grouped into rigid *planar* peptide units, there are no extra parameters required to express the three-dimensional position of the attached O and H peptide atoms. Hence, these bond lengths and bond angles are also known and fixed.

A key element of this simplified polypeptide model is that each sidechain is classified as either hydrophobic or polar, and is represented by only a single “virtual” center of mass atom. Since each sidechain is represented by only the single center of mass “virtual atom” C_s, no extra parameters are needed to define the position of each sidechain with respect to the backbone mainchain. The twenty amino acids are thus classified into two groups, hydrophobic and polar, according to the scale given by Miyazawa and Jernigan in [11].

Corresponding to this simplified polypeptide model is a simple energy function. This function includes four components: a contact energy term favoring pairwise hydrophobic residues, a second contact term favoring hydrogen bond formation between donor NH and acceptor C′=O pairs, a steric repulsive term which rejects any conformation that would permit unreasonably small interatomic distances, and a main chain torsional term that allows only certain preset values for the backbone dihedral angle pairs (φ,ψ). Since the residues in this model come in only two forms, hydrophobic and polar, where the hydrophobic monomers exhibit a strong pairwise attraction, the lowest free energy state involves those conformations with the greatest number of hydrophobic “contacts” (see [5]) and intrastrand hydrogen bonds. Despite its simplicity, the use of this type of potential function has been successful in studies by Sun, Thomas, and Dill [18] and by Srinivasan and Rose [16]. The specific potential function used initially in this study is a simple modification of the Sun/Thomas/Dill energy function and has the following form:

$$E_{total} = E_{ex} + E_{hp} + E_{hb} + E_{\varphi\psi}$$

where E_{ex} is the steric repulsive term which rejects any conformation that would permit unreasonably small interatomic distances, E_{hp} is the contact energy term favoring pairwise hydrophobic residues, E_{hb} is the contact energy term favoring pairwise hydrogen bonding, and $E_{\varphi\psi}$ is the main chain torsional term that allows only those (φ,ψ) pairs which are permitted by the Ramachandran maps. In particular, the excluded volume energy term E_{ex} and the hydrophobic interaction energy term E_{hp} are defined in this case as follows:

$$E_{ex} = \sum_{ij} \frac{C_1}{1.0 + exp((d_{ij} - d_{eff})/d_w)}, \text{ and}$$

$$E_{hp} = \sum_{|i-j|>2} \varepsilon_{ij} f(d_{ij}) \quad \text{where}$$

$$f(d_{ij}) = \frac{C_2}{1.0 + exp((d_{ij} - d_0)/d_t)}.$$

The excluded volume term E_{ex} is a soft sigmoidal potential where d_{ij} is the interatomic distance between two C_α atoms or between two sidechain center of mass atoms C_s, d_w determines the rate of decrease of E_{ex}, and d_{eff} determines the midpoint of the function (i.e. where the function equals 1/2 of its maximum value). Similarly, the hydrophobic interaction energy term E_{hp} is a short ranged soft sigmoidal potential where d_{ij} represents the interatomic distance between two sidechain centroids C_s, and d_0 and d_t represent the rate of decrease and the midpoint of E_{hp}, respectively. The hydrophobic interaction coefficient ε_{ij} = -1.0 when both residues i and j are hydrophobic, and is set to 0 otherwise. Figure 2 shows the combined effect of the energy terms $E_{ex} + E_{hp}$ for a pair of hydrophobic residues.

The contact energy term E_{hb} represents the attraction between backbone amides and carbonyls participating in a hydrogen bond. To determine this energy of attraction, it is assumed that the most favorable hydrogen bond is one in which all of the bonded pairs C′/O, O/H, and H/N lie along a straight line, since the energies between the pairs C′/H and O/N are repulsive while the energies between C′/N and O/H are attractive (see Figure 3). If for the ij^{th} pair of carbonyl and amide groups, we define $d^{(1)}_{ij}$ to be the distance between the C′ and H constituents, $d^{(2)}_{ij}$ to be the distance between C′ and N, $d^{(3)}_{ij}$ to be the distance between O and H, and $d^{(4)}_{ij}$ to be the distance between the O and N, then the energy term E_{hb} can be represented as follows (with an additional steric repulsive energy term, not shown, between the O/H pair as well):

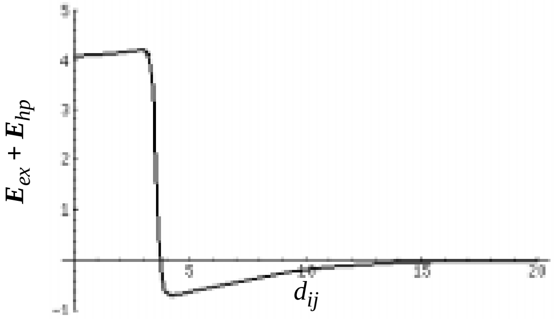

Figure 2 Combined Potential Function Energy Terms $E_{ex} + E_{hp}$

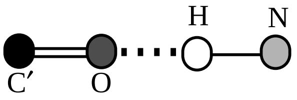

Figure 3 Schematic Illustration of Most Favorable Hydrogen Bond Alignment

$$E_{hb} = C_3 \sum_{ij} \sum_{k=1}^{4} E_{ijk},$$

where $E_{ij1} = Q_C Q_H / d^{(1)}_{ij}$, $E_{ij2} = Q_C Q_N / d^{(2)}_{ij}$, $E_{ij3} = Q_O Q_H / d^{(3)}_{ij}$, and $E_{ij4} = Q_O Q_N / d^{(4)}_{ij}$, and where Q_C = +0.5, Q_H = +0.3, Q_O = -0.5, and Q_N = -0.3 are the charges on the four participating atoms. This energy term is computed for all pairs ij of backbone amide and carbonyl groups.

The final term in the potential energy function, $E_{\varphi\psi}$, is the torsional penalty term allowing only “realistic” (φ,ψ) pairs in each conformation. That is, since φ and ψ refer to rotations of two rigid peptide units around the same C_α atom (see Figure 1), most combinations produce steric collisions either between atoms in different peptide groups or between a peptide unit and the side chain attached to C_α (except for glycine). Hence, only certain specific combinations of (φ,ψ) pairs are actually observed in practice, and are often expressed in terms of the Ramachandran plot, such as the one in Figure 4. The φ-ψ search space is therefore very restricted.

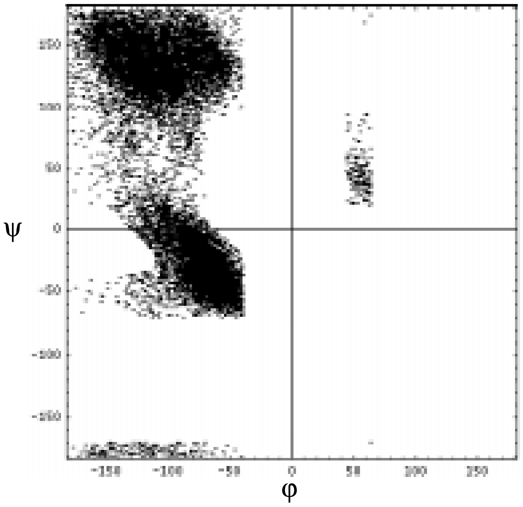

Figure 4 Typical Ramachandran Plot

The Ramachandran restrictions help maintain reasonable local conformations. To speed up the local minimizations we require a differentiable function representing the Ramachandran energies. Our approach [7] is to model the Ramachandran data by a smooth function which will have the approximate value zero in any permitted region, and a large positive value in all excluded regions.

A key observation in the construction of the function $E_{\varphi\psi}$ is that the set of allowable (φ,ψ) pairs form compact clusters in the φ-ψ plane. By enclosing each such cluster in an appropriately constructed ellipsoid, we may use the ellipsoids to define the energy term $E_{\varphi\psi}$. In particular, given p regions (ellipsoids) R_1, R_2,..., R_p, containing the experimentally allowable (φ,ψ) pairs (see Figure 5), we want the energy term $E_{\varphi\psi}$ to satisfy

$$(1) \qquad E_{\varphi\psi} \cong \begin{cases} 0 \text{ if } (\varphi, \psi) \in R_i \text{ for some } i \\ \beta \qquad\qquad \text{otherwise} \end{cases}$$

where β is some large constant penalty. To obtain such an energy term, we first represent the i^{th} ellipsoid R_i by a quadratic function $q_i(\varphi,\psi)$ which is positive definite (both eigenvalues positive) and satisfies $q_i(\varphi,\psi) = 0$ on the boundary of the ellipsoid R_i, $q_i(\varphi,\psi) < 0$ in the interior of R_i, and $q_i(\varphi,\psi) > 0$ in the exterior. By simply constructing a sigmoidal penalty term of the form

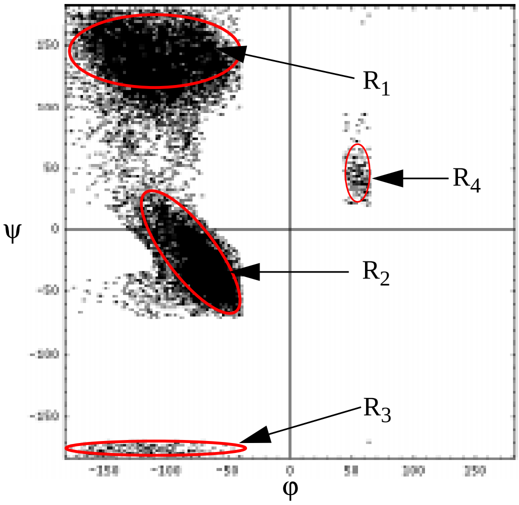

Figure 5 Approximating Ramachandran Data by Ellipsoids

$$(2) \qquad E_{\varphi\psi} = \frac{\beta}{1 + \sum_{i=1}^{p} exp(-\gamma_i q_i(\varphi, \psi))}$$

where the constants $\gamma_i > 0$ determine the rate by which $E_{\varphi\psi}$ approaches 0 or β near an ellipsoid boundary, then it is easy to see that $E_{\varphi\psi} \cong 0$ in the ellipsoid's interior, and $E_{\varphi\psi} \cong \beta$ at distant exterior points, thus satisfying (1). Figure 6 shows $E_{\varphi\psi}$ as a function of $q_i(\varphi,\psi)$ for a single ellipsoid with the values of $\beta = 1$ and $\gamma_1 = 5$. Figure 7 illustrates the three-dimensional plot of $E_{\varphi\psi}$ (for $\beta = 1$, and all $\gamma_i = 100$) for the data provided in Figure 4. This differentiable representation for $E_{\varphi\psi}$ is crucial for applying a continuous minimization technique to the problem of computing the minimum potential

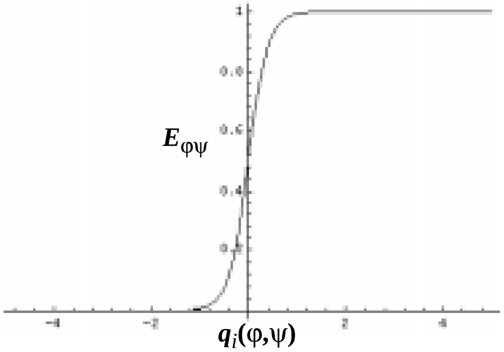

Figure 6 Sigmoidal Penalty Term $E_{\varphi\psi}$ for $\beta = 1$ and $\gamma_1 = 5$

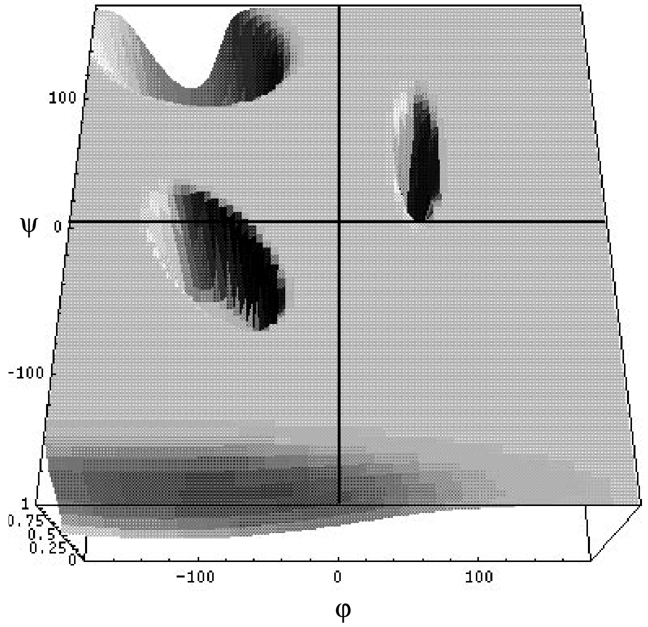

Figure 7 Three-Dimensional Plot of $E_{\varphi\psi}$ (for $\beta = 1.0$) Corresponding to the Ramachandran Data in Figure 5

energy, and its use in the CGU algorithm, described next, is a major reason for the speed of this method.

THE CGU GLOBAL OPTIMIZATION ALGORITHM

One computational method for finding the global minimum of the polypeptide's potential energy function is to use a global underestimator to localize the search in the region of the global minimum. This CGU (convex global underestimator) method, first described in [13] and subsequently applied successfully to the molecular model and potential energy function described earlier (see [6] and [7]), is designed to fit all known local minima with a convex function which underestimates all of them, but which differs from them by the minimum possible amount in the discrete L_1 norm (see Figure 8). The use of such an underestimating function allows the replacement of a *very* complex function by a simple convex underestimator. For simplicity of notation, we define the differentiable function $F(\phi) \equiv E_{total}$, where $\phi \in \mathbf{R}^{\tau}$ ($\tau = 2n\text{-}2$ represents the number of backbone dihedral angles φ and ψ), and where $F(\phi)$ is assumed to have many local minima.

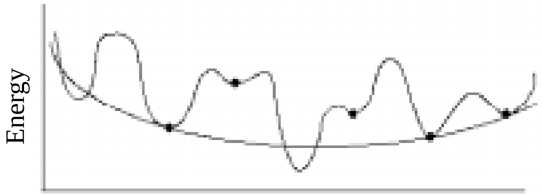

Figure 8 The Convex Global Underestimator (CGU)

To begin the iterative process, a set of $K \geq 2\tau+1$ distinct local minima $\phi^{(j)}$, for $j=1,...,K$, are computed and a convex quadratic underestimator function $\Psi(\phi)$ is then fitted to these local minima so that it underestimates all the local minima, and normally interpolates $F(\phi^{(j)})$ at $2\tau+1$ points (see Figure 8). This is accomplished by determining the coefficients in the function $\Psi(\phi)$ so that

$$\delta_j = F(\phi^{(j)}) - \Psi(\phi^{(j)}) \geq 0$$

for $j=1,...,K$, and where $\sum_{j=1}^{k} \delta_j$ is minimized. That is, the difference between $F(\phi)$ and $\Psi(\phi)$ is minimized in the discrete L_1 norm over the set of K local minima $\phi^{(j)}$, $j=1,...,K$. The underestimating function $\Psi(\phi)$ is given by

$$\Psi(\phi) = c_0 + \sum_{i=1}^{\tau}\left(c_i\phi_i + \frac{1}{2}d_i\phi_i^2\right). \tag{3}$$

Convexity of this quadratic function is guaranteed by requiring that $d_i \geq 0$ for $i=1,...,\tau$. Other linear combinations of convex functions could also be used, but the coefficients c_i and d_i of this particular quadratic function provide various useful information related to the energy landscape.

The convex quadratic underestimating function $\Psi(\phi)$ determined by the values $c \in \mathbf{R}^{\tau+1}$ and $d \in \mathbf{R}^{\tau}$ provides a global approximation to the local minima of $F(\phi)$, and its easily computed global minimum point ϕ_{min} is given by $(\phi_{min})_i = -c_i/d_i$, $i=1,...,\tau$, with corresponding function value Ψ_{min} given by

$$\Psi_{min} = c_0 - \sum_{i=1}^{\tau} c_i^2/(2d_i). \tag{4}$$

The value Ψ_{min} is a good candidate for an approximation to the global minimum of the correct energy function $F(\phi)$, and so ϕ_{min} can be used as an initial starting point around which additional configurations (i.e. local minima) can be generated. That is, the minimum of this underestimator is used to predict the global minimum for the true potential function $F(\phi)$, allowing a more localized conformer search to be performed based on the predicted minimum (see Figures 9 and 10). A new set of conformers generated by the localized search then serves as a basis for another quadratic underestimation over a sufficiently reduced space. After several repetitions, the reduced space will contain only a single conformation so that the convex underestimator prediction agrees exactly with the best known local minimum (see [13] for specific details). This conformation thus has the minimum energy among all conformations found by the algorithm, and so it is taken to be the best approximation to the true global minimum conformation.

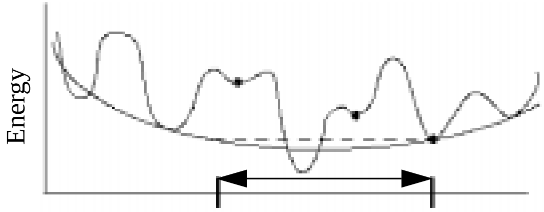

Figure 9 A Reduced Search Domain

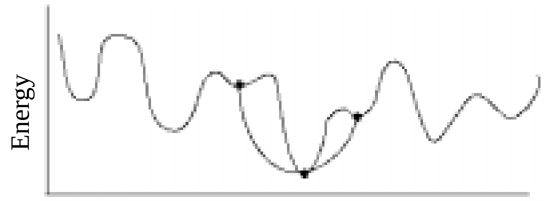

Figure 10 The New CGU Over the Reduced Search Domain

THE NATIVE STRUCTURE AND ITS FLUCTUATIONS

The CGU method gives a very simple expression for the fluctuations around the native structure. The probability $P(\phi^{(j)})$ of finding a molecule in a conformation $\phi^{(j)}$ is given by the Boltzmann distribution law

$$P(\phi^{(j)}) = \frac{e^{-(E_j - E_0)/kT}}{\sum_{i=0}^{N} e^{-(E_i - E_0)/kT}} \tag{5}$$

where $E_j \equiv F(\phi^{(j)})$ is the energy of the conformation $\phi^{(j)}$, E_0 is the energy of the global minimum conformation $\phi^{(0)}$, kT is Boltzmann's constant multiplied by temperature, and $N \geq K$ is the total number of conformations that are local minima of $F(\phi)$. Thus, higher energy states are less probable than lower energy ones.

Note that the CGU method will find only *K* of the *N* total local minima of $F(\phi)$, and that *K* « *N* is expected. However, for those *N*-*K* local minima not found, the corresponding energies are also expected to satisfy *E* » E_0, so that their effect on the total sum in (5) is negligible.

If $(\phi^{(j)})_i$ represents the i^{th} angle in the j^{th} conformation, then the weighted mean of the i^{th} angle is given by

$$\langle \phi_i \rangle = \sum_{j=0}^{N} P(\phi^{(j)}) \cdot (\phi^{(j)})_i$$

and the corresponding mean square deviation in $<\phi_i>$ is given by

$$\langle [\phi_i - \langle \phi_i \rangle]^2 \rangle = \sum_{j=0}^{N} P(\phi^{(j)}) \cdot [(\phi^{(j)})_i - \langle \phi_i \rangle]^2.$$

Thus a small mean square deviation demonstrates the increased reliability of $<\phi_i>$. Also a small mean square deviation should give $<\phi_i> \approx (\phi^{(0)})_i$. If all such mean square deviations are small, then the computed global minimum angles $(\phi^{(0)})_i$ should give a good approximation to the true native conformation.

If the final convex underestimator $\Psi(\phi)$ agrees with the global minimum potential energy E_0 at the computed global minimum conformation $\phi^{(0)}$ (this condition can be easily assured by requiring that $c_i = -d_i (\phi^{(0)})_i$ for all $i = 1,...,\tau$; that is, the gradient of the convex underestimator vanishes at the computed global minimum), then E_0 can be expressed, using (4), as

$$E_0 = \Psi_{\min} = c_0 - \sum_{i=1}^{\tau} c_i^2/(2d_i). \tag{6}$$

Furthermore, with a suitable assumption[1], combining (3) and (6) shows that the convex underestimator energy $\Psi(\phi)$ of any conformation ϕ is related to the computed global minimum energy E_0 by

[1] For each conformation $\phi^{(j)}$, the CGU function value $\Psi(\phi^{(j)})$ matches the corresponding potential energy function $E_j \equiv F(\phi^{(j)})$. Even if this assumption is not satisfied, an upper bound on the standard deviation given in (8) may be obtained.

$$\Psi(\phi) - E_0 = \frac{1}{2}\sum_{i=1}^{\tau} d_i[\phi_i - (\phi^{(0)})_i]^2.$$

Now, if $\phi_{(l)}$ denotes a conformation with all angles ϕ_i, except for ϕ_l, fixed at their respective global minimum values $(\phi^{(0)})_i$, then the energy difference directly attributed to any ϕ_l is clearly

(7) $$\Psi(\phi_{(l)}) - E_0 = \frac{1}{2} d_l[\phi_l - (\phi^{(0))})_l]^2.$$

Finally, by applying (7) to (5), the Boltzmann distribution of angle ϕ_l is then proportional to (ignoring the denominator in (5))

$$P(\phi_{(l)}) = e^{-\frac{1}{2}d_l[\phi_l - (\phi^{(0)})_l]^2/kT} = e^{\frac{-1}{2\sigma_l^2}[\phi_l - \bar{\phi}_l]^2}$$

where $\bar{\phi}_l$ is the mean, and σ_l^2 is the variance. Therefore, we can interpret $(\phi^{(0)})_l$ as the mean value of ϕ_l, and kT/d_l as the variance of ϕ_l obtained directly from the convex global underestimator. Note that a large value of d_l implies a small variance in the angle ϕ_l. Also a high temperature T, as well as a small value of d_l, implies a large variance in the angle, as expected. The standard deviation of ϕ_l is

(8) $$\sigma_l = (kT/d_l)^{1/2}.$$

Note that this result depends on the property that the CGU algorithm computes a large set of local minima, in addition to the global minimum.

COMPUTATIONAL SUMMARY

Using the chain representation, energy function, and CGU search algorithm described above, we used eight protein sequences as test cases (met-enkephalin, bradykinin, oxytocin, mellitin, zinc-finger, avian pancreatic polypeptide, crambin, and BBA1 [17], a 23-residue ββα motif). These are test cases only insofar as they provide sequences of amino acids we can model having reasonable chain lengths. The structures of some of these are not known. Our aim here is only to see if we can reach the global minimum of the mathematical model for each of these sequences. While the CGU algorithm has been extensively tested on a variety of high performance platforms including the Intel Paragon, the Cray T3D, an 8 workstation Dec Alpha cluster, and a heterogeneous network of 13 Sun SparcStations and 7 SGI Indys, the results presented here represent those obtained from the heterogeneous network of Suns and SGIs using MPI (Message Passing Interface) for the interprocess communication. The total time for solution and the computed global minimum potential energy for each structure is given in Table 1. Computations on the Cray T3D are in progress and are expected to result in substantial reductions in computing time.

The two principal results of this paper are: (1) to show that starting from many different randomly chosen open starting conformations of the chain, the CGU method converges on the same structure in each case, suggesting that the method is probably reaching the global minimum of the energy function, and (2) to show the scaling of the solution time with the chain length (see Table 1), indicating that the method seems practical for small protein-sized molecules. Table 2 shows that the simulations are dominated by a single stable state for chains longer than about 20 residues.

This paper is a test of a conformational search strategy, not an energy function. The energy function is not yet an accurate model of real proteins: the best computed structures differ from the true native structures. But similarly simple energy functions have begun to show value in predicting protein structures ([16], [18], and [22]). Therefore we believe improved energy functions used in conjunction with the CGU search method may be useful in protein folding algorithms.

ACKNOWLEDGMENTS

K.A. Dill and A.T. Phillips were supported by the NSF grant BIR-9119575, and J.B. Rosen was supported by the ARPA/NIST grant 60NANB4D1615 and NSF grants CCR-9509085 and CCR-9527151. The authors also wish to thank the San Diego Supercomputer Center for providing the computational resources to conduct this research.

REFERENCES

1. C.B. Anfinsen, *Principles that Govern the Folding of Protein Chains*, Science **181** (1973), 223-230.
2. E.M. Boczko, and C. Brooks, *First-Principles Calculation of the Folding Free Energy of a Three-Helix Bundle Protein*, Science **269** (1995), 393-396.
3. D.G. Covell, *Folding Protein α-Carbon Chains into Compact Forms by Monte Carlo Methods*, PROTEINS:

Structure, Function, and Genetics **14** (1992), 409-420.

4. D.G. Covell, *Lattice Model Simulations of Polypeptide Chain Folding*, Journal of Molecular Biology **235** (1994), 1032-1043.
5. K.A. Dill, *Dominant Forces in Protein Folding*, Biochemistry **29** (1990), 7133-7155.
6. K.A. Dill, A.T. Phillips, and J.B. Rosen, *CGU: An Algorithm for Molecular Structure Prediction*, IMA Volumes in Mathematics and its Applications (1996), forthcoming.
7. K.A. Dill, A.T. Phillips, and J.B. Rosen, *Molecular Structure Prediction by Global Optimization*, Journal of Global Optimization (1996), forthcoming.
8. D. Hinds, and M. Levitt, *Exploring Conformational Space with a Simple Lattice Model for Protein Structure*, Journal of Molecular Biology **243** (1994), 668-682.
9. I. Kuntz, G. Crippen, P. Kollman, and D. Kimmelman, *Calculation of Protein Tertiary Structure*, Journal of Molecular Biology **106** (1976), 983-994.
10. M. Levitt, and A. Warshel, *Computer Simulation of Protein Folding*, Nature **253** (1975), 694-698.
11. S. Miyazawa, and R.L. Jernigan, *A New Substitution Matrix for Protein Sequence Searches Based on Contact Frequencies in Protein Structures*, Protein Engineering **6** (1993): 267-278.
12. A. Monge, R. Friesner, and B. Honig, *An Algorithm to Generate Low-Resolution Protein Tertiary Structures from Knowledge of Secondary Structure*, Proceedings of the National Academy of Science USA **91** (1994), 5027-5029.
13. A.T. Phillips, J.B. Rosen, and V.H. Walke, *Molecular Structure Determination by Convex Global Underestimation of Local Energy Minima*, Dimacs Series in Discrete Mathematics and Theoretical Computer Science **23** (1995), P.M. Pardalos, G.-L. Xue, and D. Shalloway (Eds), 181-198.
14. M. Sippl, M. Hendlich, and P. Lackner, *Assembly of Polypeptide and Protein Backbone Conformations from Low Energy Ensembles of Short Fragments: Development of Strategies and Construction of Models for Myoglobin, Lysozyme, and Thymosin Beta 4*, Protein Science **1** (1992), 625-640.
15. J. Skolnick, and A. Kolinski, *Simulations of the Folding of a Globular Protein*, Science **250** (1990), 1121-1125.
16. R. Srinivasan and G.D. Rose, *LINUS: A Hierarchic Procedure to Predict the Fold of a Protein*, PROTEINS: Structure, Function, and Genetics **22** (1995), 81-99.
17. M.D. Struthers, R.P. Cheng, and B. Imperiali, *Design of a Monomeric 23-Residue Polypeptide with Defined Tertiary Structure*, Science **271** (1996), 342-345.
18. S. Sun, P.D. Thomas, and K.A. Dill, *A Simple Protein Folding Algorithm using a Binary Code and Secondary Structure Constraints*, Protein Engineering **8** (1995), 769-778.
19. S. Vajda, M.S. Jafri, O.U. Sezerman, and C. DeLisi, *Necessary Conditions for Avoiding Incorrect Polypeptide Folds in Conformational Search by Energy Minimization*, Biopolymers **33** (1993), 173-192.
20. A. Wallqvist, and M. Ullner, *A Simplified Amino Acid Potential for use in Structure Predictions of Proteins*, PROTEINS: Structure, Function, and Genetics **18** (1994), 267-280.
21. C. Wilson, and S. Doniach, *A Computer Model to Dynamically Simulate Protein Folding - Studies with Crambin*, PROTEINS: Structure, Function, and Genetics **6** (1989), 193-209.
22. K. Yue, and K.A. Dill, *Folding Proteins with a Simple Energy Function and Extensive Conformational Searching*, Protein Science **5** (1996), 254-261.

Table 1 Eight Small Test Problems

compound (residues)	solution time (minutes)	potential energy (kcal/mol)
met-enkephalin (5)	1.6	-44.26
bradykinin (9)	4.0	-21.89
oxytocin (9)	8.3	-121.76
BBA1 (23)	334.5	-715.51
mellitin (27)	594.5	-903.76
zinc-finger (30)	422.6	-284.64
avian pancreatic polypeptide (36)	834.3	-381.66
crambin (46)	2239.6	-734.94

Table 2 Probability Distribution of Local Minima

	Number of Local Minima in Probability Range Shown					
compound (residues)	1-.8	.8-.6	.6-.4	.4-.2	<.2	Total
met-enkephalin (5)		1		1	4	6
bradykinin (9)		1		1	22	24
oxytocin (9)	1				20	21
BBA1 (23)	1				51	52
mellitin (27)	1				67	68
zinc-finger (30)	1				143	144
avian pancreatic polypeptide (36)	1				132	133
crambin (46)	1				246	247